Here I define the Bayesian model I want to use to test the performance of the different priors. The features I want to include are the folowing: continuous outcome, any number/measurement scale.

Let's start with the **model**

$$y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\theta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i + \epsilon_{ij} \tag{1}$$

$\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi})$
$\epsilon_{ij} \sim N(0, \sigma^2)$

(Vectors are in bold, matrix are capital Greek letters).
I'm going to define the following **priors**:

$$p(\boldsymbol{\theta}) \propto 1 \tag{2}$$

$$p(\sigma^2) \propto \sigma^{-2} \tag{3}$$

For what concerns the random effects variance covariance matrix, different priors are tested. In particular we used:

- inverse-Wishart

  $$p(\boldsymbol{\Psi}) \propto IW(\nu_0, S_0) \tag{4}$$

  where we choose ν_0 to be 1 and S_0 to be diag(2), following indications by Gelman *et al* (2014).

- inverse-Wishart *a là* Huang and Wand

  $$\begin{aligned} p(\boldsymbol{\Psi}|\boldsymbol{a_1}, \boldsymbol{a_2}) &\propto IW(\nu_0 + p - 1, 2\nu_0 diag(1/a_1, 1/a_2)),\\ a_k &\propto IG(1/2, 1/A_k^2), \end{aligned} \tag{5}$$

  with $\nu_0 = 2$ and $\boldsymbol{A} = [100, 100]$. Considering a 2×2 random effects variance covariance matrix (random intercepts, and random slopes) that has this prior distribution, the corresponding marginal distributions of the standard deviation random components have an half-Cauchy distribution (belonging to the half-t family, see Gelman (2006) for details). In particular, any square-root diagonal element σ_k of a matrix following such a distribution is Half-$t(\nu_0, A_k)$. When, $\nu = 1$, and A_k is equal to the prior guess we have an half-Cauchy prior distribution for a σ_k parameter. Furthermore, the marginal priors for the correlations have a beta distribution with parameters $\alpha = -1, \beta = -1$ (Huang and Wand, 2013).

- Matrix-F variate

  $$\begin{aligned} p(\boldsymbol{\Psi}) &\propto F(\boldsymbol{\Psi}; \nu_0, \delta, \boldsymbol{B})\\ &\propto \int IW(\boldsymbol{\Psi}; \delta + k - 1, \Sigma) \times W(\boldsymbol{\Sigma}; \nu_0, \boldsymbol{B}) d\boldsymbol{\Sigma} \end{aligned} \tag{6}$$

where $\nu_0 = 2$, $\delta = 1$, and $\boldsymbol{B}$ is a prior guess. Three different choices where made for $\boldsymbol{B}$ in this paper: diag(10^3), proper neighbor of $(\sigma^2)^{-\frac{1}{2}}$; $\boldsymbol{B}_{ed}$, an educated guess based on data exploration, $\boldsymbol{R^*}$ and an empirical bayes choice following Kass and Natarajan (2006).
Considering a 2×2 random effects variance covariance matrix (random intercepts, and random slopes) that is matrix-F distirbuted, $F(\nu_0, \delta, \boldsymbol{B})$, the marginal distribution on the standard deviations of the random effects are univariate $F(\nu_0, \delta, b_{11})$ and $F(\nu_0, \delta, b_{22})$, with $\nu_0 > 1, \delta > 0, b_{jj} > 0$.

The derivation of the conditional posterior follows.

Full conditional for $\boldsymbol{\theta}$ (fixed effects)

Let's start with

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}_i, \boldsymbol{\Psi}, \sigma^2) = p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}_i, \boldsymbol{\Psi}, \sigma^2)p(\boldsymbol{\theta}) \tag{7}$$

where

$p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2) = \prod_{i=1}^n \prod_{j=1}^J p(y_{ij}|\boldsymbol{\theta}^T \boldsymbol{x}_{ij} + \boldsymbol{b}_i^T \boldsymbol{z}_{ij}, \boldsymbol{\Psi}, \sigma^2) \propto exp(-\frac{1}{2\sigma^2} SSR)$

and

$SSR = \sum_{i=1}^n [\sum_{j=1}^J (y_{ij} - \boldsymbol{\theta}^T \boldsymbol{x}_{ij} - \boldsymbol{b}_i^T \boldsymbol{z}_{ij})^2]$
where can rewrite y_{ij} as $\tilde{y}_{ij}$, with $\tilde{y}_{ij} = y_{ij} - \boldsymbol{b}_i^T \boldsymbol{z}_{ij}$. This would turn SSR in:

$SSR = \sum_{i=1}^n [\sum_{j=1}^J (\tilde{y}_{ij} - \boldsymbol{\theta}^T \boldsymbol{x}_{ij})^2] =$
$= (\tilde{\boldsymbol{y}} - \boldsymbol{X\theta})^T (\tilde{\boldsymbol{y}} - \boldsymbol{X\theta})$
$= \tilde{\boldsymbol{y}}^T \tilde{\boldsymbol{y}} - 2\boldsymbol{\theta}^T \boldsymbol{X} \tilde{\boldsymbol{y}} + \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{X\theta}$

Hence,

$$p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2) \propto exp(-\frac{1}{2\sigma^2}[-2\boldsymbol{\theta}^T \boldsymbol{X} \tilde{\boldsymbol{y}} + \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{X\theta}]) \tag{8}$$

Combining this with the prior we obtain:

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2) \propto exp(-\frac{1}{2}\boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{X\theta} + \boldsymbol{\theta}^T \boldsymbol{X} \tilde{\boldsymbol{y}}) \tag{9}$$

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2) \sim multivariate - N(\frac{(\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X} \tilde{\boldsymbol{y}}}{\sigma^2}, \frac{(\boldsymbol{X}^T \boldsymbol{X})^{-1}}{\sigma^2}) \tag{10}$$

Full conditional for $\boldsymbol{b}_i$ (random effects)
To derive this one we can start from:

$$p(\boldsymbol{b}_i|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\Psi}, \sigma^2) = p(\boldsymbol{y_i}|\boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2)p(\boldsymbol{b}_i) \tag{11}$$

We know that

$p(\boldsymbol{y}_i|.) = \prod_{j=1}^{J} p(y_{ij}|\boldsymbol{\theta}^T \boldsymbol{x}_{ij} + \boldsymbol{b}_i^T \boldsymbol{z}_{ij}, \sigma^2) \propto exp(-\frac{1}{2\sigma^2} SSR_i)$

with

$SSR = \sum_{j=1}^{J} (y_{ij} - \boldsymbol{\theta}^T \boldsymbol{x}_{ij} - \boldsymbol{b}_i^T \boldsymbol{z}_{ij})^2$
and we can rewrite y_{ij} as $\tilde{y}_{ij} = y_{ij} - \boldsymbol{\theta}^T \boldsymbol{x}_{ij}$, which would make SSR be

$SSR = \sum_{j=1}^{J} (\tilde{y}_j - \boldsymbol{\theta}^T \boldsymbol{x}_j)^2 =$
$= (\tilde{\boldsymbol{y}} - \boldsymbol{b}_i^T \boldsymbol{Z}_i)^T (\tilde{\boldsymbol{y}} - \boldsymbol{b}_i^T \boldsymbol{Z}_i)$
$= \tilde{\boldsymbol{y}}^T \tilde{\boldsymbol{y}} - 2\boldsymbol{b}_i^T \boldsymbol{Z}_j \tilde{\boldsymbol{y}} + \boldsymbol{b}_i^T \boldsymbol{Z}_i^T \boldsymbol{Z}_j \boldsymbol{b}_i$

Hence,

$$p(\boldsymbol{y}_i|.) \propto exp(-\frac{1}{2\sigma^2}[-2\boldsymbol{b}_i^T \boldsymbol{Z}_j \tilde{\boldsymbol{y}} + \boldsymbol{b}_i^T \boldsymbol{Z}_i^T \boldsymbol{Z}_j \boldsymbol{b}_i]) \tag{12}$$

We also know that in this case, the "prior" is

$$p(\boldsymbol{b}_i) \propto N(\boldsymbol{0}, \boldsymbol{\Psi}) \propto exp(-\frac{1}{2}[-2\boldsymbol{b}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{0} + \boldsymbol{b}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{b}_i]) \tag{13}$$

In conclusion, combining the sampling model and the prior, we get:

$$p(\boldsymbol{b}_i|.) \propto exp(-\frac{1}{2\sigma^2}[-2\boldsymbol{b}_i^T \boldsymbol{Z}_j \tilde{\boldsymbol{y}} + \boldsymbol{b}_i^T \boldsymbol{Z}_i^T \boldsymbol{Z}_j \boldsymbol{b}_i] - \frac{1}{2\sigma^2}[-2\boldsymbol{b}_i^T \boldsymbol{Z}_j \tilde{\boldsymbol{y}} + \boldsymbol{b}_i^T \boldsymbol{Z}_i^T \boldsymbol{Z}_j \boldsymbol{b}_i]) \tag{14}$$

$$p(\boldsymbol{b}_i|.) \propto multiv\!-\!N((\Psi^{-1} + \frac{\boldsymbol{Z}_i^T \boldsymbol{Z}_i}{\sigma^2})^{-1}(\boldsymbol{\Psi}^{-1}\boldsymbol{0} + \frac{\boldsymbol{Z}_i^T \tilde{y}_i}{\sigma^2}), (\Psi^{-1} + \frac{\boldsymbol{Z}_i^T \boldsymbol{Z}_i}{\sigma^2})^{-1}) \tag{15}$$

Full conditional for σ^2 (error variance) The full conditional posterior can be expressed as:

$$p(\sigma^2|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{\Psi}) = p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2) p(\sigma^2) \tag{16}$$

The sampling model is the same we saw for the full conditional distribution of $\boldsymbol{\theta}$:

$p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2) = \prod_{i=1}^{n} \prod_{j=1}^{J} p(y_{ij}|\boldsymbol{\theta}^T \boldsymbol{x}_{ij} + \boldsymbol{b}_i^T \boldsymbol{z}_{ij}, \boldsymbol{\Psi}, \sigma^2) =$
$= \prod_{i=1}^{n} \prod_{j=1}^{J} (2\pi\sigma^{-2})^{-\frac{1}{2}} exp(-\frac{(y_{ij} - \boldsymbol{\theta}^T \boldsymbol{x}_{ij} - \boldsymbol{b}_i^T \boldsymbol{z}_{ij})^2}{2\sigma^2})$

However, we are now interested in σ^2, hence

$p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Psi}, \sigma^2) \propto (\sigma^2)^{-\frac{N}{2}} exp(-\frac{\sum_{i=1}^{n} \sum_{j=1}^{J} (y_{ij} - \boldsymbol{\theta}^T \boldsymbol{x}_{ij} - \boldsymbol{b}_i^T \boldsymbol{z}_{ij})^2}{2\sigma^2}) = (\sigma^2)^{-\frac{N}{2}} exp(-\frac{1}{2\sigma^2} SSR)$

where $N = \sum_i^n nj_i$ is the entire sample size (all observations within all clusters).

The prior for σ is given above, and therefore we can write the full conditional posterior as:

$$p(\sigma^2|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{\Psi}) \propto (\sigma^2)^{-\frac{N}{2}-1} exp(-\frac{1}{2\sigma^2} SSR) \tag{17}$$

$$p(\sigma^2|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{b}_i, \boldsymbol{\Psi}) \sim IG(\frac{N}{2}, \frac{SSR}{2}) \tag{18}$$

Full conditional for $\boldsymbol{\Psi}$ (random effects variance covariance matrix)

Here, we need to write down the posteriors for the different priors we specified. First, let us define the sampling model for the random effects.
$M = {}^{a\ b}_{c\ d}$

$$\begin{gathered} \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} = \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\Psi}) \\ p(\boldsymbol{b}_1, \boldsymbol{b}_2|\boldsymbol{\Psi}) \propto |\boldsymbol{\Psi}|^{-\frac{n}{2}} \end{gathered} \tag{19}$$

- inverse-Wishart
- inverse-Wishart *a là* Huang and Wand
- Matrix-F variate

See Mulder Pericchi, 2018

Notation Conventions

- n number of clusters; i specific cluster
- J number of observations within cluster; j specific observation
- N total number of observations